Statistical mechanics of optimization problems

Giorgio Parisi

Dipartimento di Fisica, INFM, SMC and INFN,

Università di Roma *La Sapienza*, P. A. Moro 2, 00185 Rome, Italy.

Abstract

Here I will present an introduction to the results that have been recently obtained in constraint optimization of random problems using statistical mechanics techniques. After presenting the general results, in order to simplify the presentation I will describe in details the problems related to the coloring of a random graph.

I. INTRODUCTION

In statistical mechanics [1] the partition function is

$$Z(\beta) = \sum_{\mathcal{C}} \exp(-\beta H(\mathcal{C})) \;, \tag{1}$$

where $\beta = 1/(kT)$ and $\mathcal{C}$ is a generic configuration.

In optimization problems [2] we want to find the configuration $\mathcal{C}^*$ that minimizes the function $H(\mathcal{C})$ and to know the minimal cost, $H^* \equiv H(\mathcal{C}^*)$. We define $E(T)$ and $S(T)$ to be respectively the expectation value of the energy and the entropy as function of the temperature T. H^* is $E(0)$ and the number of optimizing configurations is $\exp(S(0))$. In order to obtain information on the nearly optimizing configurations, i.e. those configurations such that $H(\mathcal{C}) = H^* + \epsilon$, we must study the system at small temperature. We are interested to knowing what happens in the thermodynamic limit, i.e. when the number N of variables goes to infinity.

In optimization theory it is natural to consider an ensemble of Hamiltonians and to find out the properties of a generic Hamiltonian of this ensemble [3,4], in the same way as in

the theory of disordered systems. Sometimes the ensemble is defined in a loose way, e,g. problems that arise from practical instance such as chips placements on computer boards. In other words we have an Hamiltonian that is characterized by a set of parameters (denoted by J) and we have a probability distribution $\mu(J)$ on this parameter space. We want compute the ensemble average

$$\int d\mu(J) E_J(T) \equiv \overline{E_J(T)} \,. \tag{2}$$

We are interested in computing the probability distribution $P(E)$ of the zero temperature energy E over the ensemble. When the number N of variables goes to infinity, if E is well normalized, and its probability distribution becomes a delta function [5–7]: intensive quantity do not fluctuate in the thermodynamic limit.

II. CONSTRAINT OPTIMIZATION

In a typical case a configuration of our system is composed by N variables σ_i that take q values (e.g. from 1 to q). A instance of the problems is characterized by M functions $f_k[\sigma]$ ($k = 1, M$), each function takes only the values 0 or 1.

Let us consider the following example with $N = 4$ and $M = 2$:

$$c_1[\sigma] = \theta(\sigma_1\sigma_2 - \sigma_3\sigma_4) \,, \quad c_2[\sigma] = \theta(\sigma_1\sigma_3 - \sigma_2\sigma_4) \,. \tag{3}$$

The function we want to minimize is

$$H[\sigma] = c_1[\sigma] + c_2[\sigma]. \tag{4}$$

We are interested to know if there is a minimum with $H[\sigma] = 0$. If this happens all the function c_k must be zero. The condition $H[\sigma] = 0$ is equivalent to the following two inequalities:

$$\sigma_3\sigma_4 > \sigma_1\sigma_2 \quad , \sigma_2\sigma_4 > \sigma_1\sigma_3 \,. \tag{5}$$

Each function imposes a constraint: the function H is zero if all the constraints are satisfied: we are in the satisfiable case. If not possible to satisfy all the constraints, the minimal total energy is different from zero and we stay in the unsatisfiable case.

Given N and M we define the ensemble as all the possible different sets of M inequalities of the type

$$\sigma_{i_1(k)}\sigma_{i_2(k)} > \sigma_{i_3(k)}\sigma_{i_4(k)} \ . \tag{6}$$

The interesting limit is when N goes to infinity with

$$M = N\alpha \ , \tag{7}$$

α being a parameter. Hand waving arguments suggest that for small α it is should be possible to satisfy all the constraints, while for very large α most of the constraints will be not satisfied. We define the energy density

$$e(\alpha) = \lim_{N\to\infty} \frac{\overline{H^*}}{N} \ . \tag{8}$$

There is a phase transition at a critical value of α_c, such that

$$e(\alpha) = 0 \quad \text{for} \quad \alpha <= \alpha_c \ , \quad e(\alpha) > 0 \quad \text{for} \quad \alpha > \alpha_c \ . \tag{9}$$

III. RANDOM GRAPHS AND BETHE APPROXIMATION

We define a random Poisson graph [8] in the following way: given N nodes we consider the ensemble of *all* possible graphs with $M = \alpha N$ edges (or links). A random Poisson graph is a generic element of this ensemble. The local coordination number z_i is the number of nodes that are connected to the node i. The average coordination number z is given by $z = 2\alpha$.

These graphs are locally a tree: if we take a generic point i, the subgraph composed by those points that are at a distance less than d on the graph is a tree with probability one when N goes to infinity. If $z > 1$ the nodes percolate and a finite fraction of the graph belongs to a single giant connected component. Loops do exist on this graph, but they have typically a length proportional to $\ln(N)$. The absence of small loops is crucial because we can study the problem locally on a tree and we have eventually to take care of the large loops as self-consistent boundary conditions at infinity.

Random graphs are sometimes called Bethe lattices, because a spin model on such a graph has the moral duty to be soluble exactly using the Bethe approximation. Let us recall the Bethe approximation for the two dimensional Ising model [9,10]. In the standard mean field approximation, one arrives to the well known equation

$$m = \text{th}(\beta J z m) \; , \tag{10}$$

where $z = 4$ on a square lattice ($z = 2d$ in d dimensions) and J is the spin coupling. The critical point is $\beta_c = 1/z$. This result is not very exciting in two dimensions (where $\beta_c \approx .44$) and it is very bad in one dimensions (where $\beta_c = \infty$). The aim of Bethe was to obtain a better results still keeping the simplicity of the mean field theory.

Let us consider the system where a spin σ has been removed. There is a cavity in the system and the spins τ are on the border of this cavity. We assume that these spins are uncorrelated and they have a magnetization m_C. When we add the spin σ, we find that the probability distribution of this spin is proportional to

$$\sum_{\tau_i} \exp \left(\beta J \sigma \sum_{i=1,4} \tau_i \right) \prod_{i=1,4} P_{m_C}[\tau_i]) \; . \tag{11}$$

The magnetization of the spin σ is thus

$$m = \text{th}\{z \; \text{arth}[\text{th}(\beta J) m_C]\} \; , \tag{12}$$

with $z = 4$.

Now we remove one of the spin τ_i and form a larger cavity (two spins removed). We assume that the spins on the border of the cavity are uncorrelated and they have the same magnetization m_C. We obtain

$$m_C = \text{th}\{(z-1)\text{arth}[\text{th}(\beta J) m_C]\} \; . \tag{13}$$

Solving this last equation we can find the value of m_C and using the previous equation we can find the value of m. It is rather satisfactory that in 1 dimensions ($z = 2$) the cavity equations become

$$m_C = \text{th}(\beta J) m_C \; . \tag{14}$$

This equation for finite β has no non-zero solutions, as it should be. The internal energy and the free energy can also be computed. The result is not exact because the cavity spins are correlated.

If we remove a node of a random lattice [11,12], the nearby nodes (that were at distance 2 before) are now at a very large distance, i.e. $O(\ln(N))$ with probability one. In this case we hope that we can write

$$\langle \tau_{i_1} \tau_{i_2} \rangle \approx m_{i_1} m_{i_2} \ . \tag{15}$$

This happens in the ferromagnetic case in presence if a in infinitesimal of magnetic field where the magnetization may take only one value. In more complex cases, (e.g. antiferromagnets) there are many different possible values of the magnetization because there are many equilibrium states. The cavity equations become equations for the probability distribution of the magnetizations. This case have been long studied in the literature and we say that the replica symmetry is spontaneously broken [13,14]. Fortunately for the aims of this talk we need only a very simple form of replica symmetry breaking and we are not going to describe the general formalism.

IV. COLORING A GRAPH

For a given graph G we would like to know if using q colors the graph can be colored in such a way that adjacent nodes have different colors [15]. The Hamiltonian is

$$H_G = \sum_{i,k} A(i,k) \delta_{\sigma_i, \sigma_i} \ , \tag{16}$$

where $A_G(i,k)$ is the adjacency matrix and the variables σ may take values that go from 1 to q. This Hamiltonian describes the antiferromagnetic Potts model with q states. For large N on a random graph energy density does not depend on G:

$$e(z) = 0 \quad \text{for} \quad z < 1 \ , \quad e(z) \propto \sqrt{z} \quad \text{for} \quad z \to \infty \ . \tag{17}$$

There is a phase transition at z_c between the colorable phase $e(z) = 0$ and the uncolorable phase $e(z) \neq 0$. For $q = 2$ we have $z_c = 1$: Odd loops cannot be colored and for $z > 1$

there are many large loops that are even or odd with equal probability. The $q = 2$ case is an antiferromagnetic Ising model on a random graph, i.e. a standard spin glass.

Let us consider a legal coloring (i.e all adjacent nodes have different colors). We take a node i and we consider the subgraph of nodes at distance d from a given node. Let us call $\mathcal{B}(i, d)$ the interior of this graph. We ask the following questions:

- Are there other legal colorings of the graph that coincide with the original coloring outside $\mathcal{B}(i, d)$ and differs inside $\mathcal{B}(i, d)$? We call the set of all these coloring $\mathcal{C}(i, d)$.

- Which is the list of colors that the node i may have in one of the coloring belonging to $\mathcal{C}(i, d)$? We call this list $L(i, d)$. This list depends on the legal coloring we started from.

$L(i, d)$ has a limit when d goes to infinity. We call this limit $L(i)$, i.e. the list of all the possible colors that the site i may have if we change only the colors of the nearby nodes and we do not change the colors of faraway nodes.

Let us study what happens on a graph where the site i has been removed. We denote by k a node adjacent to i and we call $L(k; i)$ the list of the possible colors of the node k. The various nodes k do not interact directly and their colors are independent.

In this situation it is evident that $L(i)$ can be written as function of all the $L(k; i)$. We have to consider all the neighbors (k) of the node i; if a neighbor may be colored in two ways, it imposes no constraint, if it can be colored in only one way, it forbids the node i to have its color. Considering all nearby nodes we construct the list of the forbidden colors and the allowed colors are those colors that are not forbidden.

A further simplification may be obtained if we associate to a list L a variable ω, that take values from 0 to q, defined as follow

- The variable ω is equal to i if the list contains only the ith color.

- The variable ω is equal to 0 if the list contains more than one color.

In the nutshell we have introduced an extra color, white. A site is white if it can be colored in more than two ways without changing the colors of the far away sites [16,17].

6

The equations are just the generalization of the Bethe equation where we have the colors, white included, instead of the magnetizations. We have discrete, not continuos variables, because we are interested in the ground state, not in the behavior at finite temperature. The previous equation are called the *belief* equations or TAP equations. We can associate to any legal coloring a solution of the belief equations. Sometimes the solution of the belief equations is called a *whitening*, because some nodes that where colored in the starting legal configuration becomes white.

Each legal coloring has many other legal colorings nearby that differs only by the change of the colors of a *small* number of nodes. The number of these legal coloring that can be reached starting from a given coloring by making this kind of moves is usually exponentially large and correspond to the same whitening.

We have three possibilities.

- For all the legal configurations the corresponding whitenings have all nodes white.

- For a generic legal configurations the corresponding whitening is non-trivial. i.e. for a finite fraction of the nodes are not white.

- The graph is not colorable and there are no legal configurations.

In the second case we want to know the number of whitenings , how they differs and which are their properties, e.g. how many sites are colored. In this case the set of all the legal configurations breaks in an large number of different disconnected regions that are called with many different names (states, valleys, clusters, lumps...). Each whitening is associated to a different cluster of legal solutions [16].

We consider the case where there is a large number of non-equivalent whitening We introduce the probability $P_i(c)$ that for a generic whitening we have that $\omega(i) = c$. The quantities $P_i(c)$ generalize the physical concept of magnetization. We will assume that for points i and l that are far away on the graph the probability $P_{i,l}(c_1, c_2)$ factorizes into the product of two independent probabilities $P_{i,l}(c_1, c_2) = P_i(c_1)P_l(c_2)$. This hypothesis in not innocent: there are many cases where it is not correct.

A similar construction can be done with the cavity coloring and in this way we define the probabilities $P_{i;k}(c)$, where k is a neighbor of i. These probabilities are called **surveys** [18,19,23]. Under the previous hypothesis the surveys satisfy equations (the so called survey propagation equations) that are simple, but are lengthy to be written. The survey propagation equations always have a trivial solution corresponding to all sites white: $P_i(0) = 1$ for all i. Depending on the graph there can be also non-trivial solutions of the survey equations. Let us assume that if such a solution exist, it is unique.

We are near the end of our trip. If we consider the whole graph we can define the probability $\mathcal{P}[P]$, i.e. the probability that a given node has a probability $P(c)$. With some work one arrives to an integral equation for $\mathcal{P}[P]$, i.e. the probabilities of the surveys, whose solution can be easily computed numerically on present days computers. One finds that there is a range $z_d < z < z_U$ where the previous integral equation has a non-trivial solution and its properties can be computed.

In the same way that the entropy counts the number of legal colorings, the complexity counts the number of different whitening; more precisely for a given graph we write

$$\#\text{whitenings} = \exp(\Sigma_G) , \tag{18}$$

where Σ_G is the complexity.

There is a simple way to compute the complexity. It mimics the standard computation of the free energy and it consists in counting the variation in the number of whitenings when we modify the graph. At the end of the day we find the results shown in fig.(1).
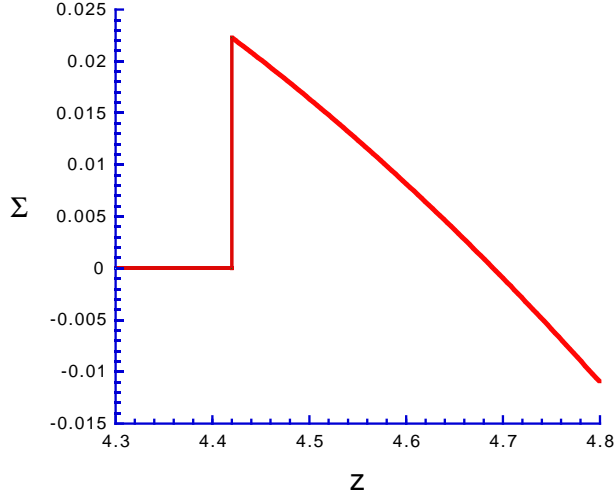
FIGURES

FIG. 1. The complexity versus the average connectivity z for three colors.

he complexity jumps from 0 to a finite value at $z_d = 4.42$; it decreases with increasing z and becomes eventually negative at $z = 4.69$. A negative value of Σ implies a number of whitenings less than one and it is interpreted as the signal there there are no whitening (and no legal configurations). In the region where the complexity is negative a correct computation of the energy $e(z)$ gives a non-zero (positive) result. The value where the complexity becomes zero is thus identified as the colorability threshold $z_c = 4.69$. Similar results may be obtained for higher values of q [15]. The concept of complexity emerged in the study of the spin glasses and it was also introduced in the study of glasses under the name of configurational entropy. The behavior of the complexity as function of z is very similar to what is supposed to happen in glasses as function of β [20,21].

V. OPEN PROBLEMS

There are many problems that are still open:

- The extension to other models.

- Verification of the self-consistency of the different hypothesis.

- The construction of effective algorithms for finding a solution of the optimization problems. A first algorithm has been proposed and it has been later improved by adding backtracking [23,24]. A goal is to produce an algorithm that for large N finds a solution on a random graph in a polynomial time as soon as $z < z_c$. Finding this algorithm is interesting from the theoretical point of view (it is not clear at all if such an algorithm does exist) and it may have practical applications.

- One should be able to transform the results derived in this way into rigorous theorems. After a very long effort Talagrand [25], using some crucial results of Guerra [26], has been recently able to prove that a similar, but more complex, construction gives the correct results in the case of infinite range spin glasses, i.e. the Sherrington Kirkpatrick model, that was the starting point of the whole approach. Some of these results have been extended to the case of the Bethe Lattice [27].

REFERENCES

[1] See for example: Parisi G. Statistical Field Theory (Academic Press, New York) 1987.

[2] Martin O. C., Monasson R. and Zecchina R., *Theoretical Computer Science* **265** (2001) 2.

[3] G.Parisi *Constraint Optimization and Statistical Mechanics*, cond-mat/0301157 (2003).

[4] Garey M. R. and Johnson D. S., *Computers and intractability* (Freeman, New York) 1979.

[5] Dubois O. Monasson R., Selman B. and Zecchina R., *Phase Transitions in Combinatorial Problems*, *Theoret. Comp. Sci.* **265**, (2001), G. Biroli, S. Cocco, R. Monasson, *Physica A* **306**, (2002) 381.

[6] Kirkpatrick S. and Selman B., *Critical Behaviour in the satisfiability of random Boolean expressions*, *Science* **264**, (1994) 1297.

[7] Dubois O., Boufkhad Y., Mandler J., *Typical random 3-SAT formulae and the satisfiability threshold*, in *Proc. 11th ACM-SIAM Symp. on Discrete Algorithms*.

[8] P. Erdös and A. Rènyi, Publ. Math. (Debrecen) **6**, 290 (1959).

[9] Thouless D.J., Anderson P.A. and Palmer R. G., *Phil. Mag.* **35**, (1977) 593.

[10] Katsura S., Inawashiro S. and Fujiki S., *Physica* **99A** (1979) 193.

[11] Mézard M. and Parisi G.. *Eur.Phys. J.* B **20** (2001) 217.

[12] Mézard M. and Parisi G.. *J. Stat. Phys* **111**, (2003) 1 .

[13] Mézard, M., Parisi, G. and Virasoro, M.A. *Spin Glass Theory and Beyond*, (World Scientific, Singapore) 1997.

[14] Parisi G., *Field Theory, Disorder and Simulations*, (World Scientific, Singapore) 1992.

[15] Mulet R., Pagnani A., Weigt M., Zecchina R., *Phys. Rev. Lett.* **89**, 268701 (2002);

Braunstein A., Mulet R., Pagnani A., Weigt M., Zecchina R., *Phys. Rev. E* **68**, (2003) 036702.

[16] Parisi G.. cs.CC/0212047 *On local equilibrium equations for clustering states* (2002).

[17] Parisi G., *On the probabilistic approach to the random satisfiability problem* cs.CC/0308010 (2003).

[18] Mézard M., Parisi G. and Zecchina R., *Science* **297**, (2002) 812.

[19] Mézard M.and Zecchina R. *Phys. Rev. E* **66**, 056126 (2002).

[20] Parisi G., *Glasses, replicas and all that* cond-mat/0301157 (2003).

[21] Cugliandolo T.F., *Dynamics of glassy systems* cond-mat/0210312 (2002).

[22] Parisi. G. *On the survey-propagation equations for the random K-satisfiability problem* cs.CC/0212009 (2002).

[23] Parisi. G. *Some remarks on the survey decimation algorithm for K-satisfiability* cs.CC/0301015 (2003).

[24] Parisi G. *A backtracking survey propagation algorithm for K-satisfiability*, cond-mat/0308510 (2003).

[25] Talagrand M. *Spin Glasses. A challenge for mathematicians. Mean-field models and cavity method*, (Springer-Verlag Berlin) 2003, *The Parisi formula*.

[26] Guerra F., *Comm. Math. Phys.* **233** (2002) 1; Guerra F. and Toninelli F.L., *Comm. Math. Phys.* **230** (2002), 71.

[27] S. Franz and M. Leone, *J. Stat. Phys* **111** (2003) 535.